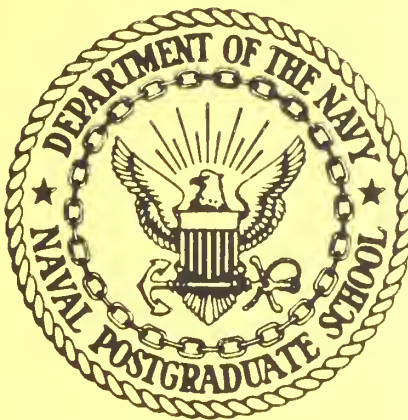


423
NPS55-87-013

NAVAL POSTGRADUATE SCHOOL

Monterey, California



SUGGESTED IMPROVEMENTS FOR THE JOINT
OIL ANALYSIS PROGRAM CORRELATION PROGRAM

HAROLD J. LARSON
TOKE JAYACHANDRAN

OCTOBER 1987

Approved for public release; distribution unlimited.

Prepared for:
Kelly AFB
TX 79241

FedDocs
D 208.14/2
NPS-55-87-013

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral R. C. Austin
Superintendent

K. T. Marshall
Acting Provost

The work reported herein was supported in part with funds from
Kelly AFB.

Reproduction of all or part of this report is authorized.

This report was prepared by:

REPORT DOCUMENTATION PAGE

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY CA 93943-5101

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b RESTRICTIVE MARKINGS	
2a SECURITY CLASSIFICATION AUTHORITY		3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b DECLASSIFICATION / DOWNGRADING SCHEDULE		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPS55-87-013		7a NAME OF MONITORING ORGANIZATION	
5a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b OFFICE SYMBOL (If applicable) Code 55	7b. ADDRESS (City, State, and ZIP Code)	
5c ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000		9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8a NAME OF FUNDING / SPONSORING ORGANIZATION Wright AFB, Texas	8b OFFICE SYMBOL (If applicable) SA-ALC/MMEI	10. SOURCE OF FUNDING NUMBERS	
8c ADDRESS (City, State, and ZIP Code) Wright AFB, Texas 79241		PROGRAM ELEMENT NO.	PROJECT NO. 86-717
		TASK NO.	WORK UNIT ACCESSION NO.

TITLE (Include Security Classification)

SUGGESTED IMPROVEMENTS FOR THE JOINT OIL ANALYSIS PROGRAM CORRELATION PROGRAM

PERSONAL AUTHOR(S)

Larson, Harold J. and Jayachandran, Toke

11. TYPE OF REPORT Technical	13b TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) 1987, Oct	15 PAGE COUNT 29
---------------------------------	---	--	---------------------

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Interlaboratory testing, spectrometric oil analysis, correlation program.
FIELD	GROUP	SUB-GROUP	

ABSTRACT (Continue on reverse if necessary and identify by block number)

The Joint Oil Analysis Correlation Program is intended to provide a monthly indication of how consistently the various Department of Defense (DOD) oil analysis laboratories are performing. The procedure employed is a modification of one suggested by Youden in 1959; the modification employed does not appear to provide a fair measure for all the types of spectrometers used by DOD. This paper provides a comparison of the current procedure results with those which would occur with Youden's original procedure, and suggests that the latter approach is preferable.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Harold J. Larson		22b TELEPHONE (Include Area Code) (408)646-2473	22c OFFICE SYMBOL Code 55La

INTRODUCTION

The Joint Oil Analysis Program (JOAP) coordinates the Department of Defense (DOD) programs employing spectrometric analyses of used oils for condition monitoring of many types of equipment. Two hundred-odd different oil analysis laboratories provide these spectrometric analyses, the great majority of which are individual-service owned, with the remainder being contract facilities (non DOD). Because of the mobility of equipment, it is quite possible that successive samples of used oil, from the same piece of equipment, may not be analysed on the same instrument. For this reason, and numerous others, it is highly desirable that the same oil sample, when analyzed by different laboratories, should as nearly as possible result in the same contaminant readings.

In the mid 1970's JOAP instituted their "correlation" program, intended to provide information regarding the consistency of readings produced by the spectrometric instruments serving their needs; this program was expected to monitor both internal consistency of repeated readings by the same instrument, as well as consistency from one instrument (or laboratory) to another. The landmark paper discussing this type of problem is by Youden [3], which highlights some empirical observations about instrument-to-instrument testing in general. The procedure Youden describes for checking laboratory to laboratory consistency consists of sending each participating laboratory two "similar" samples of unknown composition; each laboratory receives the same two samples. Each is required to analyze both of the samples (one time) and return the pair of results to a central processing location. If one defines

x_i = Analysis result for sample 1, laboratory i

y_i = Analysis result for sample 2, laboratory i

then the n pairs (x_i, y_i) , $i = 1, 2, \dots, n$, can be represented as n points in a plane. If one plots these n points, Youden pointed out that the resulting swarm of points almost invariably has the general shape depicted in Figure 1. With coordinate axes at the medians (or means) of the x_i and y_i values (as in Figure 2), the preponderance of points will typically fall in the first and third quadrants, with relatively few in the second and fourth. This would necessarily follow in a situation in which a laboratory tends to get either high readings or low readings, for both of the two samples; if we were to draw in a 45° line and project the points onto this line, the resulting scatter of these projected points describes laboratory-to-laboratory variation. This variation Youden attributed to differences in laboratory technique (or could equally well

Figure 1. Plot of Sample 1 versus Sample 2

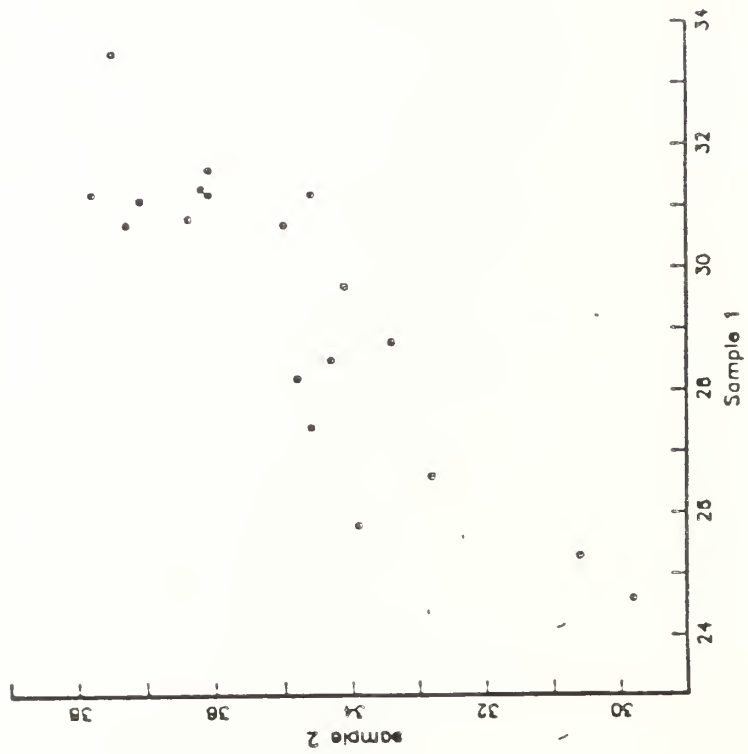
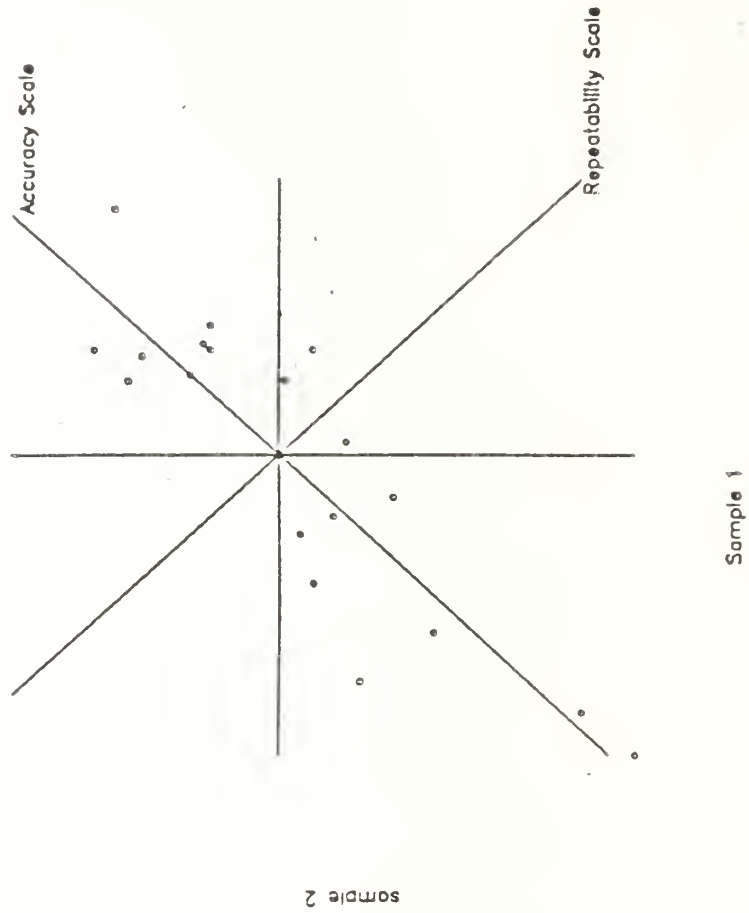


Figure 2. Accuracy and Repeatability Scales



be called the variation in accuracy of the laboratories). Youden also pointed out that one can measure the perpendicular distance of each point from this 45° line (i. e., also project the original points onto the normal to the 45° line) to measure the "precision" (or repeatability) of a given laboratory; scatter in this direction should be mainly due to the ability of an individual laboratory to reproduce its own results. Youden suggested that limits defining acceptable laboratory performance can be constructed from the scatter or variation observed in these two directions.

THE CURRENT CORRELATION PROGRAM

The JOAP correlation program was modelled after the interlaboratory type of comparison described by Youden, with some important modifications; the basic computations used in the JOAP correlation program are described in [1]. This program is administered by the JOAP Technical Support Center (TSC), located at the Naval Air Rework Facility, Pensacola, FL. Briefly, the correlation program works as follows: Each JOAP laboratory is sent the same pair of oil samples (actually 2 pairs of samples are used, as described later), each month; the particular concentrations of the elements of interest in these samples vary from month to month and are not known by the participating laboratories. Each laboratory analyzes the pair of samples it receives (presumably only once) and mails the results back to the TSC. Again, let (x_i, y_i) represent the two sample readings from laboratory i , for a given element. The procedure described in [1] first determines a "trimmed" mean value for the x 's and for the y 's, independently. These trimmed means are computed by arranging the given x analyses, say, in order of magnitude, deleting the lowest 20% and the highest 20%, and then averaging the remaining middle 60%. Note that it is quite possible that the x score from laboratory 1 might be trimmed off, while its y score is not; that is, a given laboratory's results may contribute to one trimmed mean and not the other. Note as well that only 60% of the x scores received, and 60% of the y scores, are used to define these trimmed means. Letting x_T and y_T represent these trimmed means (for a given element and month), the JOAP correlation procedure locates a new coordinate system at (x_T, y_T) ; these trimmed means play the role of the medians in Youden's discussion [3].

Rather than constructing a 45° line, as suggested by Youden, reference [1] uses a line of slope S , where S is determined by the trimmed means (x_T, y_T) and constants A_j, B_j which differ from element to element and are presented in Table 1.

Table 1. Constants used to determine slope

Element	A	B
Fe	2.0	.1
Ag	1.5	.1
Al	2.0	.1
Cr	1.5	.1
Cu	1.5	.1
Mg	1.5	.1
Si	1.9	.14
Ti	1.5	.1
Ni	1.5	.1

The slope S used for a given element is defined by

$$S = (A^2 + B^2 y_T^2)^{.5} / (A^2 + B^2 x_T^2)^{.5}.$$

If $x_T = y_T$, this formula gives $S = 1$ (an angle of 45°); indeed, if the two samples sent to a given laboratory have essentially equal concentrations of a given element (the guidelines call for the two to differ by no more than 15%), then this computed slope will not differ from 1 by a great deal. Reference [1] says this formula is intended to avoid "an error unless the composition of the material being measured is identical in the two samples". It is not clear what this expected error might have been, nor does it appear that the slope used by the procedure will materially differ from 1 as Youden suggested. The pair of readings (x_i, y_i) are then projected onto the line with slope S (giving the accuracy score for the laboratory) and onto the line which is normal to this line with slope S (giving the repeatability score for the laboratory).

The major way in which the computations for the JOAP correlation program differ from the procedure suggested by Youden is in the manner in which accuracy (laboratory variation) and repeatability (variability of repeated analyses by the same laboratory with the same sample) are assessed. Reference [1] mentions a current (1973) laboratory certification program which was designed to assure that each laboratory could meet minimum standard performance criteria. This certification program calls for the laboratory to conduct a sequence of ten separate analyses of a prepared oil standard with known concentration c , say, of a given element. The accuracy index (AI) of the laboratory for

this element is the magnitude of the difference between the known concentration, c , and the average of the laboratory's 10 analyses; the repeatability index (RI) of the laboratory for this element is the sample standard deviation of the 10 analyses, computed in the usual way. The acceptable limit for AI is

$$M = (A^2 + B^2 c^2)^{.5}$$

where the A and B values are those given in Table 1 above, for the specific element. The laboratory passes the accuracy certification for this element so long as $AI < M$; it passes the repeatability certification so long as $RI < M/2$. Thus the constants given in Table 1 were initially proposed for this certification program, and were undoubtedly derived from some physical model of the way in which a particular type of instrument should behave, based on ten repeated analyses of the same sample.

In the correlation program, the accuracy criterion for a given element is defined to be

$$a = (2A^2 + B^2(x_T^2 + y_T^2))^{.5}$$

and the repeatability criterion is $a/2$, where the constants A and B again come from Table 1 above. Note that a is in fact the square root of the sum of the squares of the M values for the two samples, with the trimmed means x_T , y_T playing the roles of the known concentrations c . It is curious that these same constants should be used in the correlation program, where each of two different samples is to be analyzed one time, not ten, and presumably any type of instrument might be used. Each laboratory then is judged on its accuracy and repeatability performance for each element (each month). If the magnitude of its accuracy score exceeds a , it fails on accuracy, and if the magnitude of its repeatability score exceeds $a/2$, it fails repeatability. This way of defining acceptable limits for the two types of scores depends only on the trimmed means (and the constants A and B) and in no way on the actual scatter of the observed data themselves, contrary to Youden's suggestion. It also leads to quite erratic behavior, in a certain sense, which will be explored below.

It was mentioned earlier that the correlation program actually sends two pairs of samples to each laboratory each month. One pair of samples is prepared by the TSC in new oil, using organo-metallic concentrates with added sulfonate;

it is possible to control the contamination levels of all elements of interest fairly well with these samples. This pair of samples is referred to as "synthesized" samples. In addition to the pair of synthesized samples, the TSC also sends each laboratory a pair of used engine oil samples. These are made from used contaminated oils and, as such, should behave more like actual oil samples the laboratories are expected to analyze daily. It is much more difficult for the TSC to exert control over the contaminant levels in these samples; frequently the same powdered metallic contaminants used for the synthesized samples are added to the used oil samples to adjust the contaminant levels. Thus the correlation program monitors the laboratory performances on both types of samples.

A second dichotomy exists in the correlation program, defined by the physical principle employed by the instrument in measuring concentration. Roughly 80% of the instruments used in JOAP are atomic emission (AE) spectrometers. In these instruments the sample material (the oil) is excited by an electric spark and the spectral lines of the light emitted are used to measure concentrations. The remaining 20% of the instruments used are atomic absorption (AA) spectrometers. In these instruments the sample material is excited by a gas flame, while illuminated by a light of known composition; the amount of the known light absorbed, at specific spectral lines, is used to determine the concentrations in the sample. Because of these different physical bases for measurement, it is well known that the resulting concentration scales are not identical. The correlation program computations are carried out separately for these two types of instrument. Thus a typical JOAP correlation program report contains two major partitions: one describing the behavior of the AE instruments and the other describing the behavior of the AA instruments. Within each of these two, the behaviors for synthesized oil samples and for used oil samples are examined separately, computing the trimmed means, projecting the readings onto "accuracy" and "repeatability" axes, etc., for each element of interest. Although not mentioned in [1], it is undoubtedly true that the constants A and B in table 1, used in defining the accuracy limit a , are derived from a theoretical model of the behavior of a particular atomic emission instrument; nevertheless, the same constants are employed with the AA instruments. At the present time, the same 9 elements are monitored for both types of instrument: iron (Fe), silver (Ag), aluminum (Al), chromium (Cr), copper (Cu), magnesium (Mg), silicon (Si), titanium (Ti), and nickel (Ni).

The correlation program summarizes the monthly behavior of each participating JOAP instrument by a single score, combining behavior over the synthesized and used oil samples. This score is arrived at by subtracting from 100 a certain number of points for each element that the instrument fails to pass (because of its accuracy result or its repeatability result or both) for each sample type for each month. Table 2 presents the number of points lost for each element.

Table 2. Number of points lost for failing accuracy and/or repeatability, either sample type.

Element	Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Ni
Points	9	6	4	4	9	4	4	6	4

Thus, if laboratory 1, say, had failed accuracy for Fe, both accuracy and repeatability for Cr, with synthetic samples, and only failed Cr with used samples, its monthly score would be $100 - 9 - 4 - 4 = 83$. If laboratory 2 failed accuracy for Si and Ti, synthetic samples, and both accuracy and repeatability for Cu, used samples, its score would be $100 - 4 - 6 - 9 = 81$. If a laboratory fails either accuracy or repeatability for every element, for both types of samples, notice its score would be 0. These monthly scores are used in the correlation program to track laboratory performance over time. The laboratories' 6 month average score is computed and used for certification of the laboratory. If this 6 month average score is below 80 for three consecutive months, the laboratory may be decertified; if the 6 month average score lies between 80 and 90 for 3 consecutive months, the laboratory is provisionally certified. For all other cases the laboratory is continued to be certified.

As mentioned earlier, the acceptable limits for accuracy scores and repeatability scores depend only on the trimmed means x_T , y_T and the appropriate constants from Table 1; they do not depend on the actual scatter of the accuracy or repeatability scores themselves. This causes both the accuracy and the repeatability limits, which define the acceptable values, to jump around a great deal, in terms of the number of standard deviations they represent (away from the means, which are 0). Tables 3, 4, 5, and 6 illustrate this phenomenon for the correlation data collected for August, 1986. These tables summarize the number of instruments of the two types that submitted analysis re-

Table 3. August, 1986, summary of correlation scores.

	Correlation Scores Used Oils			Atomic Emission August 1986			Number of Labs
	Limit	Accuracy #StDev	#Fail	Repeatability Limit	#StDev	#Fail	
Fe	3.1405	3.2027	10	1.5703	3.8917	0	183
Ag	2.1213	17.2925	0	1.0607	8.5170	0	183
Al	2.8284	52.0000	3	1.4142	26.0000	0	183
Cr	2.1215	3.2357	0	1.0607	6.4067	0	183
Cu	2.5010	3.1894	11	1.2505	3.8719	0	183
Mg	2.7776	1.9751	32	1.3888	1.9253	3	183
Si	2.8164	3.4913	5	1.4082	3.1023	6	183
Ti	2.1236	3.0166	1	1.0618	4.9184	0	183
Mo	2.1291	2.5732	10	1.0646	2.5786	1	182
Ni	2.1228	3.0808	1	1.0614	4.4233	0	183

Table 4. August, 1986, summary of correlation scores.

	Correlation Scores Synthetic Oils			Atomic Emission August 1986			Number of Labs
	Limit	Accuracy #StDev	#Fail	Repeatability Limit	#StDev	#Fail	
Fe	5.5883	3.9480	4	2.7941	3.0941	0	183
Ag	3.0275	2.1700	12	1.5137	1.7183	3	183
Al	5.9450	3.7681	7	2.9725	3.2787	0	183
Cr	2.9595	2.9511	4	1.4798	3.2800	1	183
Cu	3.8884	2.4030	10	1.9442	1.4166	5	183
Mg	8.5668	2.8151	20	4.2834	2.2899	4	183
Si	15.9339	4.8487	5	7.9670	4.6366	0	183
Ti	3.8749	2.6859	9	1.9374	2.8417	1	183
Mo	2.6922	1.8139	9	1.3461	2.5420	1	182
Ni	3.2006	3.6200	2	1.6003	3.1763	0	183

Table 5. August, 1986, summary of correlation scores.

Correlation Scores Used Oils			Atomic Absorption August 1986			
Limit	Accuracy #StDev	#Fail	Repeatability Limit	#StDev	#Fail	Number of Labs
2.9089	2.3724	6	1.4545	3.8233	0	37
2.1213	10.5623	1	1.0607	5.2812	0	37
2.8292	4.0772	0	1.4146	8.4949	0	37
2.1213	3.5942	0	1.0607	∞	0	37
2.2341	3.6256	0	1.1171	3.3077	0	37
2.2976	2.2609	7	1.1488	3.3621	0	37
2.7267	2.6933	11	1.3634	8.6240	3	31
2.1216	3.5411	5	1.0608	∞	0	32
2.1213	∞	1	1.0607	∞	1	14
2.1220	3.1295	5	1.0610	∞	1	35

Table 6. August, 1986, summary of correlation scores.

Correlation Scores Synthetic Oils			Atomic Absorption August 1986			
Limit	Accuracy #StDev	#Fail	Repeatability Limit	#StDev	#Fail	Number of Labs
5.4000	2.2351	7	2.7000	3.1609	0	37
2.9638	2.1089	8	1.4819	3.3577	1	37
5.5213	2.0836	7	2.7607	2.7949	2	37
2.7500	1.6746	13	1.3750	3.1476	1	37
3.6054	3.5904	4	1.8027	3.9123	1	37
8.9353	1.9332	14	4.4676	3.7263	3	37
15.1153	2.3767	11	7.5577	2.9521	0	31
3.8108	1.9218	12	1.9054	2.6810	0	32
2.7003	1.6592	4	1.3501	3.0514	1	14
3.0897	2.2816	9	1.5448	3.0922	2	35

sults for the various elements, as well as the computed acceptable limits for accuracy and repeatability, for both types of samples. In addition, the actual standard deviations of the scores have been computed. Recall that 40% of the data was trimmed, for both the x (Sample one) scores and for the y (Sample two) scores, in locating the origin for the accuracy and repeatability axes. The standard deviations used in this discussion are computed from the readings provided by those instruments which were kept after the trimming, for one or both of the two samples. In every case, the means for these values are essentially 0, so the standard deviations were computed about 0. The column labelled #StdDev gives the ratio of the limit (given in the column labelled Limit) for the given variable divided by this computed standard deviation. Note that the number of standard deviations which the limits represent vary quite widely from element to element, especially for the used samples (for both types of instrument). They also vary quite widely from month to month; Appendix A presents the same type of data for one additional month, January, 1985, the only other month for which we have all the necessary data available in electronic form for AE instruments. For AA instruments, an additional 18 months of data has been available to us; although not included with this report, there is tremendous variation in the ratio of the limit divided by the standard deviation from element to element, and from month to month for the same element. There does not seem to be any logical reason that one would like this type of ratio to vary in this way. It would seem to indicate that there is a wide variability in the ease with which a laboratory could meet the accuracy and repeatability requirements from month to month.

The instruments read out concentration values to the nearest .1; these values are reported to the TSC. The data entered into the computer for the correlation program computations is rounded to the closest integer, causing a large number of pairs of analyses to be identical. This phenomenon in turn can lead to all of the nontrimmed accuracy scores (or equally well the repeatability scores), which are used to compute the standard deviation, being equal; such a standard deviation then is 0 and the ratio of the computed limit to such a standard deviation is of course undefined. This situation is labelled by the symbol ∞ in the #StdDev column (see e. g., Table 5, Cr, repeatability). If the data were entered with full accuracy (including tenths) it is expected this phenomenon will not occur very frequently.

SUGGESTED IMPROVEMENTS

The paper by Youden, discussed earlier, led to the publication of a number of additional contributions to the literature about interlaboratory comparisons. An interesting paper [2] was published by Mandel and Lashof, giving interpretations and more mathematical discussions of Youden's ideas, one year after [1] discussed the JOAP correlation program as now implemented. Among other things, Mandel and Lashof give some models which make the scatter of points mentioned by Youden seem natural, as well as changes to these models which could reasonably lead to quite different looking plots. They suggest that the bivariate normal distribution provides a good model for the original pairs of sample readings (x_i, y_i) ; if one lets \bar{x} , \bar{y} represent the means of the observed pairs, then the pairs $(x_i - \bar{x}, y_i - \bar{y})$ will be bivariate random variables with means equal to zero. One can then use principal components to find the direction of the axis which includes the greatest variability; for some simple reasonable types of structures this direction turns out to be the line with slope 1, the phenomenon pointed out by Youden. The orthogonal direction is the one with the least variability, and is free of effects of different instruments under a standard type of linear model; for the types of samples used in the JOAP correlation program, it would appear that the simple type of linear model they discuss should be appropriate. The following discussion incorporates some of the ideas and suggestions made by Mandel and Lashof.

For a given month, for a given element, and type of instrument, let (x_i, y_i) represent the observed pairs of analyses received, $i = 1, 2, \dots, n$, where n is the number of instruments. Let us assume that

$$x_i = \mu_1 + L_i + e_i,$$

$$y_i = \mu_2 + L_i + f_i,$$

for $i = 1, 2, \dots, n$. μ_1 and μ_2 represent the "true" contents of samples 1 and 2, respectively; L_i is meant to represent a laboratory effect that is constant for both x_i and y_i , the two sample readings received from the same laboratory. e_i and f_i represent independent random measurement errors (or noise) for the two analyses. It seems quite reasonable to assume that the e_i and f_i values are independent and normally distributed with the same variance; as Mandel and Lashof suggest, one can also assume that the laboratory effects, L_i , are normally distri-

buted. It follows then that

$$X_i = x_i - \bar{x} = L_i - \bar{L} + e_i - \bar{e}$$

$$Y_i = y_i - \bar{y} = L_i - \bar{L} + f_i - \bar{f}$$

that is, the pairs (X_i, Y_i) do not depend on the true contents μ_1, μ_2 , but only on the laboratory effects and the measurement errors. The projection of (X_i, Y_i) onto the 45° line, times the square root of two, then is

$$X_i + Y_i = 2(L_i - \bar{L}) + (e_i - \bar{e}) + (f_i - \bar{f}).$$

Note that this sum is affected by the laboratory effects as well as the measurement errors. The projection of (X_i, Y_i) onto the line normal to the 45° line, times the square root of two, is

$$X_i - Y_i = (e_i - \bar{e}) - (f_i - \bar{f}),$$

a quantity which depends only on the measurement errors, and not the laboratory effects (or the true contents). With this simple type of additive model it indeed turns out that the projections on the 45° line give a reasonable idea of accuracy, or spread, among the different laboratories and the projections on the orthogonal axis depend only on the measurement errors, or repeatability, of an instrument's readings.

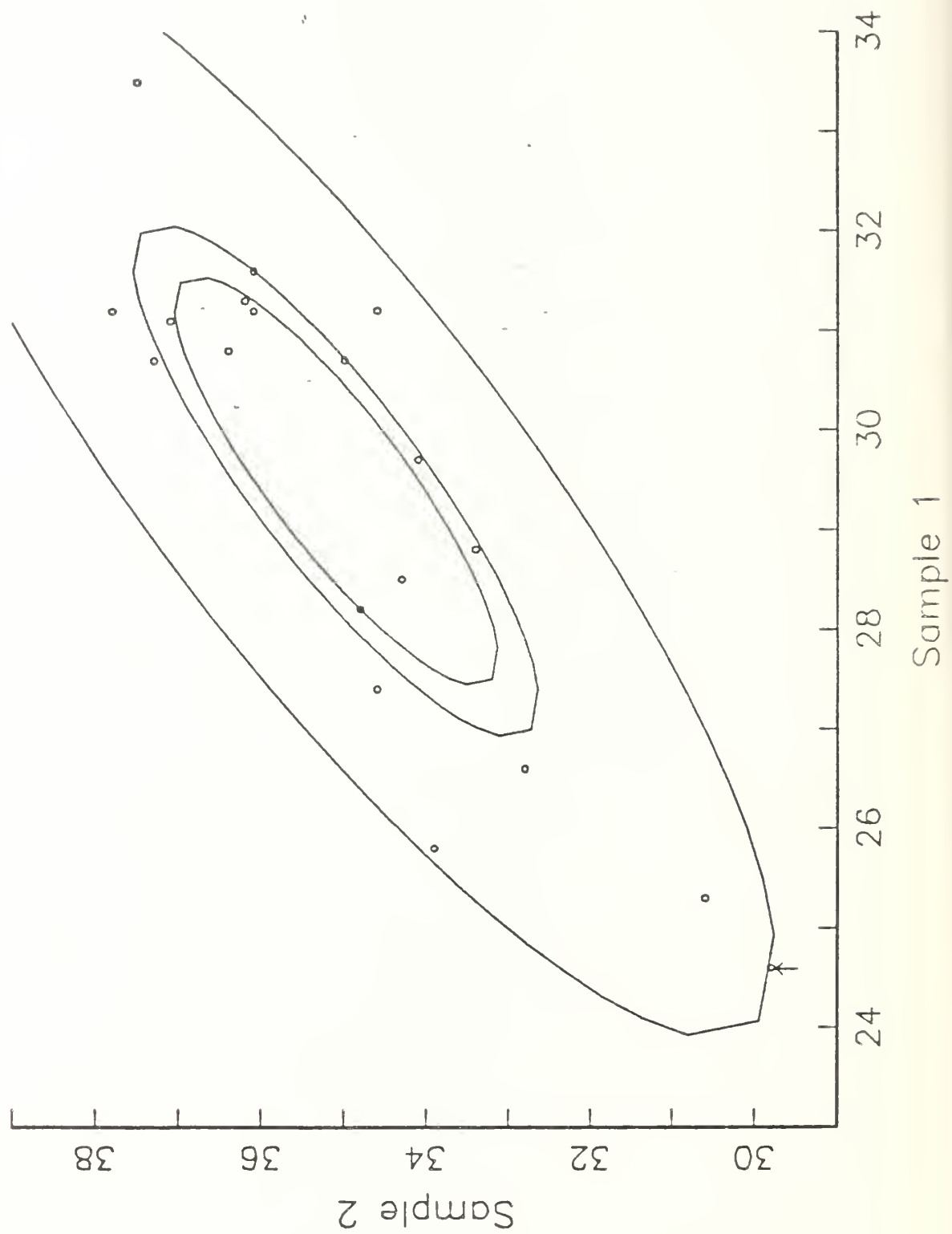
It is not uncommon for "wild" points to occur in using sensitive instruments to make fine measurements; undoubtedly the reason for trimming the data in the correlation program is to remove these effects. While we agree with this general principle (using trimming to remove outliers) it also seems that 40% trimming is very extreme. The idea of independent trimming of the two samples is also not particularly appealing, allowing, as already mentioned, the possibility that an instrument's y-score is trimmed, but its x-score is not. We have made a preliminary investigation into trimming, possible methods for doing it as well as the amount of trimming to apply. It has been pointed out some time ago that the untrimmed means and the trimmed means, using actual correlation data, result in essentially the same value for all elements for all months. That is, because of the number of laboratories participating, and the apparent fact that "wild" points seem to occur symmetricly (some big, some small), the location is

essentially unchanged if one uses untrimmed means instead of trimmed means. Since the current correlation program computations depend only on the trimmed means (and constants), one would get the same scores and results if one used the full set of raw data with no trimming. If, however, one wants to use the observed scatter or spread on the accuracy and repeatability axes to determine limits for acceptable behavior, "wild" points could seriously inflate the results; thus we are in favor of applying some trimming before establishing limits for accuracy and repeatability. We are also in favor of the philosophy of bivariate trimming: if a laboratory is trimmed on the x-scale it is also necessarily trimmed on the y-scale.

There are many different ways to implement bivariate trimming. If one adopts the suggestions of normality put forth by Mandel and Lashof, it would seem natural to use the constant contours of the bivariate normal density function to accomplish the trimming. Letting (X_i, Y_i) be as defined above, this means computing and inverting a 2 by 2 matrix (details are given in Appendix B), and then evaluating quadratic forms (locating the contours which contain the observed points). Those points most distant from the origin are the candidates for trimming. Figure 3 presents a typical scatter of observed sample results, with 3 of the bivariate normal constant contours drawn in. We have applied this type of trimming to the August, 1986, and January, 1985, data available to us; 20%, 10% and 5% trimming were looked at and for these data it appears that 5% contour trimming does a sufficiently good job for both types of instrument. Thus for the 20 points pictured in Figure 3, 5% trimming would delete the single point on the outermost ellipse.

We recommend that for a period of time (6 months or more) the scores for the correlation program be computed as present and compared with scores developed essentially according to Youden's original suggested procedure for inter-laboratory comparisons. Specifically, this second set of scores will employ 5% bivariate trimming (as mentioned above and defined in Appendix B) to guard against "wild" points. The trimmed means \bar{x} , \bar{y} are computed from the remaining observations and used to define $X_i = x_i - \bar{x}$ and $Y_i = y_i - \bar{y}$, centering at these trimmed means. Then all observed pairs (X_i, Y_i) will be projected onto the 45° line (accuracy scores) and onto the line normal to this line (repeatability scores). The original untrimmed values are then used to compute standard deviations in each of these two dimensions (see Appendix B), which in turn are used to define the acceptable accuracy and repeatability limits. Any laboratory which has an

Figure 3. Constant Bivariate Normal Contours



accuracy score whose magnitude exceeds the accuracy limit fails on accuracy; any laboratory which has a repeatability score whose magnitude exceeds the repeatability limit fails on repeatability.

The number of standard deviations to use in defining the accuracy and repeatability limits is, of course, arbitrary, and can be set at any level desired. We recommend that 3 standard deviations be used at least initially; thus the accuracy limit will be $3s_a$ and the repeatability limit will be $3s_r$, where s_a and s_r are the computed standard deviations. A rationale for using 3 as a multiplier is given below. This procedure has been applied to both the August, 1986, and the January, 1985 data; tables 7, 8, 9, and 10 present the resulting limits for the January, 1986 data and are comparable to tables 3, ..., 6 presented earlier. The computations for January, 1985, are also presented in Appendix A.

In comparing the correlation scores with these proposed scores, perhaps the most apparent difference is the increased number of laboratories which fail on repeatability (and relatively fewer on accuracy). It would appear that the correlation program method of determining the limits for repeatability (which depend only on the 60% trimmed means) does not provide an effective check. Note as well that the AA instruments in general fare much better with the proposed method than they do with the current correlation procedure. As mentioned above, the procedure for determining the limits in the correlation program includes the constants A and B which were undoubtedly derived for an AE instrument and do not perform well for an AA instrument. It is also of interest to compare which particular instruments fail on accuracy and/or repeatability for the correlation program versus the proposed procedure. Appendix A presents this information for the August, 1986, and January, 1985 data.

One rationale for determining the number of standard deviations to use in setting the accuracy and repeatability limits can be defined in terms of the chances of an instrument, which performs correctly, passing both the accuracy and repeatability tests, for all elements, for both used and synthetic samples. Let p represent the probability that a correctly functioning instrument will fail the check, for either accuracy or repeatability (the same value for both). Then the probability that it will pass both accuracy and repeatability, for any element, is $(1-p)^2$ and the probability it will pass for all nine elements for both used and synthetic samples (assuming independence) is $(1-p)^{36}$. Suppose we set this quantity equal to .9; this gives the value for p to be $1 - .9^{1/36}$, or

Table 7. Proposed method for determining scores.

	Proposed Scores Used Oils			Atomic Emission August 1986			Number of Labs
	StDev	Accuracy Limit	#Fail	StDev	Repeatability Limit	#Fail	
Fe	1.2649	3.7947	4	.3772	1.1316	2	183
Ag	.4575	1.3725	0	.0000	.0000	5	183
Al	.3116	.9347	14	.0000	.0000	15	183
Cr	.6408	1.9223	0	.1054	.3162	10	183
Cu	1.1871	3.5613	6	.2963	.8888	0	183
Mg	1.8946	5.6838	8	.3643	1.0928	3	183
Si	.9489	2.8467	5	.3752	1.1257	8	183
Ti	.7609	2.2826	1	.1571	.4712	18	183
Mo	1.0697	3.2091	0	.3541	1.0623	1	182
Ni	.6876	2.0629	2	.2243	.6729	22	183

Table 8. Proposed method for determining scores.

	Proposed Scores Synthetic Oils			Atomic Emission August 1986			Number of Labs
	StDev	Accuracy Limit	#Fail	StDev	Repeatability Limit	#Fail	
Fe	1.9422	5.8265	2	.7871	2.3614	2	183
Ag	1.4674	4.4022	3	.3993	1.1978	6	183
Al	2.2865	6.8594	7	.8169	2.4506	1	183
Cr	1.0171	3.0512	3	.4124	1.2371	2	183
Cu	1.6348	4.9044	5	.5399	1.6197	5	183
Mg	4.4506	13.3518	2	1.6082	4.8246	2	183
Si	4.8085	14.4255	5	1.5389	4.6168	2	183
Ti	1.6961	5.0882	4	.5924	1.7772	1	183
Mo	1.4966	4.4898	1	.4855	1.4564	1	182
Ni	1.0884	3.2651	2	.4284	1.2851	5	183

Table 9. Proposed method for determining scores.

Proposed Scores Used Oils			Atomic Absorption August 1986			
StDev	Accuracy Limit	#Fail	Repeatability StDev	Limit	#Fail	Number of Labs
2.1078	6.3234	0	.3357	1.0071	0	37
.4452	1.3356	1	.2226	.6678	0	37
.8396	2.5188	0	.1621	.4863	2	37
.5902	1.7706	0	.0000	.0000	0	37
.7071	2.1213	0	.3285	.9856	0	37
2.5244	7.5731	0	.4008	1.2025	0	37
3.1192	9.3576	1	.4348	1.3045	3	31
2.6074	7.8221	1	.1739	.5217	2	32
.9449	2.8347	1	.5669	1.7008	1	14
1.1114	3.3343	2	.0000	.0000	1	35

Table 10. Proposed method for determining scores.

Proposed Scores Synthetic Oils			Atomic Absorption August 1986			
StDev	Accuracy Limit	#Fail	Repeatability StDev	Limit	#Fail	Number of Labs
4.1044	12.3132	1	.8962	2.6886	0	37
2.5229	7.5687	1	.4309	1.2928	1	37
5.7968	17.3905	2	.9968	2.9905	0	37
3.0909	9.2727	2	.5297	1.5890	1	37
2.0240	6.0719	2	.6409	1.9226	1	37
19.1620	57.4860	2	1.8851	5.6553	2	37
35.6186	106.8559	0	2.3484	7.0452	1	31
5.8034	17.4102	2	.7199	2.1597	0	32
3.3033	9.9100	0	3.6025	10.8075	1	14
3.9572	11.8716	2	.7932	2.3797	2	35

.9971. With the assumption of normality for the measurement errors for the instruments, and for the variation between instruments, we then require the number of standard deviations that a normal random variable will exceed with probability .00145 (half the value of $1 - .9971 = .0029$, since the projected scores can be extreme either positively or negatively). This, in turn, results in a requirement of 2.978 standard deviations, which we have rounded to 3 for a trial of the proposed system.

BIBLIOGRAPHY

- [1] Clark, W. E., Jr., "A Procedure for Interlaboratory Correlation of Fluid Analysis Spectrometers", U. S. Naval Weapon Systems Analysis Office, Quantico, Va., 22134, WSAO-TM-731, 1973.
- [2] Mandel, John, and Lashof, T. W., "Interpretation and Generalization of Youden's Two-Sample Diagram", Journal of Quality Technology, Vol. 6, No. 1, 1974.
- [3] Youden, W. J., "Graphical Diagnosis of Interlaboratory Test Results", Industrial Quality Control, Vol. 15, No. 11, 1959.

This appendix presents additional numerical tables; Tables A1 through A8 present the correlation program data and the proposed program data for the month of January, 1985; the format and information presented is identical to that given earlier in Tables 3 through 10, for August 1986.

Table A1. January, 1985, summary of correlation scores.

	Correlation Scores Used Oils			Atomic Emission January 1985			Number of Labs
	Limit	Accuracy #StDev	#Fail	Limit	Repeatability #StDev	#Fail	
Fe	3.5071	2.6565	9	1.7535	1.9128	1	180
Ag	2.1213	3.6428	2	1.0607	9.9782	0	180
Al	2.8943	2.0591	28	1.4472	3.1567	2	180
Cr	2.1260	9.1720	1	1.0630	4.6267	0	180
Cu	3.9245	2.2462	31	1.9623	2.9665	4	180
Mg	2.1508	3.0324	5	1.0754	4.2906	0	179
Si	2.7349	3.0866	6	1.3675	2.3458	2	180
Ti	2.1260	9.9737	3	1.0630	4.9869	1	180
Mo	2.1306	2.6244	17	1.0653	1.9102	7	176
Ni	2.1243	3.0390	3	1.0622	3.3791	2	180

Table A2. January, 1985, summary of correlation scores.

	Correlation Scores Synthetic Oils			Atomic Emission January 1985			Number of Labs
	Limit	Accuracy #StDev	#Fail	Limit	Repeatability #StDev	#Fail	
Fe	10.1637	3.9005	6	5.0819	3.0955	3	180
Ag	2.6760	2.6029	19	1.3380	3.1013	1	180
Al	3.6902	2.4123	12	1.8451	2.2952	7	180
Cr	2.8611	3.5102	2	1.4305	2.4633	4	180
Cu	2.8887	2.9225	5	1.4444	3.0815	0	180
Mg	4.8431	3.0231	16	2.4215	2.3116	4	179
Si	26.1873	4.1619	8	13.0937	3.8006	0	180
Ti	3.0830	3.0546	19	1.5415	2.5627	4	180
Mo	2.1234	3.0510	2	1.0617	2.4013	5	176
Ni	2.5477	3.6203	2	1.2739	2.8681	0	180

Table A3. January, 1985, summary of correlation scores.

	Correlation Scores Used Oils			Atomic Absorption January 1985			Number of Labs
	Limit	Accuracy #StDev	#Fail	Limit	Repeatability #StDev	#Fail	
Fe	3.0387	3.2953	4	1.5194	3.9041	0	40
Ag	2.1213	17.4929	0	1.0607	8.7464	0	39
Al	2.9019	4.3052	2	1.4510	3.6814	0	39
Cr	2.1260	16.4682	1	1.0630	8.2341	0	39
Cu	2.8149	2.5119	1	1.4074	2.6220	1	40
Mg	2.1299	2.9297	2	1.0650	3.6705	0	39
Si	2.7107	1.7149	6	1.3554	2.2365	2	23
Ti	2.1214	3.9403	3	1.0607	∞	0	27
Mo	2.1213	∞	2	1.0607	∞	0	22
Ni	2.1216	3.1765	0	1.0608	8.7475	0	34

Table A4. January, 1985, summary of correlation scores.

	Correlation Scores Synthetic Oils			Atomic Absorption January 1985			Number of Labs
	Limit	Accuracy #StDev	#Fail	Limit	Repeatability #StDev	#Fail	
Fe	10.4337	2.1215	7	5.2169	2.1496	3	40
Ag	2.8148	1.5639	8	1.4074	2.8423	0	39
Al	3.5187	3.7979	4	1.7594	4.4148	0	39
Cr	2.8514	2.0918	12	1.4257	2.4828	1	39
Cu	2.8320	2.8357	2	1.4160	2.3404	2	40
Mg	4.9521	1.8625	8	2.4760	2.3912	1	39
Si	26.9246	2.5849	2	13.4623	3.0578	1	23
Ti	3.1099	1.1944	10	1.5549	2.8637	2	27
Mo	2.1260	∞	1	1.0630	∞	0	22
Ni	2.5315	2.0737	4	1.2657	1.0353	2	34

Table A5. January, 1985, summary of proposed scores.

	Proposed Scores Used Oils			Atomic Emission January 1985			Number of Labs
	StDev	Accuracy Limit	#Fail	Repeatability StDev	Limit	#Fail	
Fe	1.4253	4.2760	6	.4200	1.2600	6	180
Ag	.5860	1.7579	2	.0000	.0000	4	180
Al	2.0972	6.2915	4	.4005	1.2014	5	180
Cr	.7603	2.2810	1	.1646	.4938	14	180
Cu	2.7733	8.3198	4	.5965	1.7896	4	180
Mg	.9294	2.7882	5	.2163	.6489	22	179
Si	.8833	2.6500	6	.3574	1.0723	2	180
Ti	.7462	2.2386	3	.1405	.4216	14	180
Mo	1.1451	3.4352	0	.4113	1.2339	7	176
Ni	.7389	2.2168	3	.2340	.7021	12	180

Table A6. January, 1985, summary of proposed scores.

	Proposed Scores Synthetic Oils			Atomic Emission January 1985			Number of Labs
	StDev	Accuracy Limit	#Fail	Repeatability StDev	Limit	#Fail	
Fe	3.8243	11.4729	5	1.5377	4.6130	5	180
Ag	1.2390	3.7169	3	.3786	1.1358	2	180
Al	1.8333	5.4999	2	.6537	1.9612	7	180
Cr	.9654	2.8963	2	.5081	1.5242	4	180
Cu	1.1571	3.4712	1	.4341	1.3022	0	180
Mg	2.3905	7.1716	6	.9018	2.7054	3	179
Si	9.0119	27.0357	7	3.1042	9.3125	6	180
Ti	1.3609	4.0826	6	.5680	1.7039	1	180
Mo	.7933	2.3798	2	.3221	.9662	5	176
Ni	.8372	2.5117	2	.3957	1.1870	4	180

Table A7. January, 1985, summary of proposed scores.

	Proposed Scores Used Oils			Atomic Absorption January 1985			Number of Labs
	StDev	Accuracy Limit	#Fail	Repeatability StDev	Limit	#Fail	
Fe	1.6809	5.0427	2	.3023	.9069	1	40
Ag	.4845	1.4534	0	.0000	.0000	1	39
Al	1.6989	5.0966	0	.3288	.9864	1	39
Cr	.7076	2.1228	2	.1600	.4800	2	39
Cu	1.1350	3.4049	0	.2786	.8357	1	40
Mg	.8625	2.5876	1	.2822	.8467	0	39
Si	6.0414	18.1243	2	1.0756	3.2268	1	23
Ti	1.0539	3.1618	2	.0000	.0000	3	27
Mo	.6761	2.0284	2	.0000	.0000	2	22
Ni	.6770	2.0310	0	.0000	.0000	1	34

Table A8. January, 1985, summary of proposed scores.

	Proposed Scores Synthetic Oils			Atomic Absorption January 1985			Number of Labs
	StDev	Accuracy Limit	#Fail	Repeatability StDev	Limit	#Fail	
Fe	8.4609	25.3826	3	2.6683	8.0049	0	40
Ag	2.3520	7.0561	1	.5102	1.5306	0	39
Al	2.3069	6.9207	2	.4255	1.2764	1	39
Cr	2.7603	8.2808	0	.6728	2.0185	0	39
Cu	1.1571	3.4713	0	.5078	1.5233	2	40
Mg	5.4262	16.2786	2	1.1346	3.4038	0	39
Si	18.5882	55.7646	2	5.1979	15.5938	0	23
Ti	5.0776	15.2329	1	.8288	2.4864	0	27
Mo	.6761	2.0284	1	.0000	.0000	1	22
Ni	2.2082	6.6245	1	.5303	1.5910	1	34

Tables A9 through A24 present the indices of those laboratories which failed accuracy and/or repeatability, for the current correlation program and for the proposed method of scoring.

Table A9. Indices of Failing Labs
Correlation Scores Atomic Emission
Used Oils August 1986

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
9		11		49	1 24	9 6	182	9 12	146
33		49		55	3 181	112 33		11	
39		151		118	9 182	134 59		15	
55				146	13	155 79		41	
96				147	15	165 112		49	
152				148	16	165		78	
172				151	24			102	
177				157	33			118	
181				164	39			166	
182				181	40			180	
				182	55				
					79				
					88				
					90				
					95				
					96				
					98				
					112				
					117				
					127				
					131				
					147				
					148				
					152				
					169				
					170				
					171				
					172				
					175				
					177				
					181				
					182				

Table A10. Indices of Failing Labs
Proposed Scores Atomic Emission
Used Oils August 1986

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
152 122	2	1 1	33	55	9 24	9 6	182 42	12	146 35
177 181	3	7 7	41	118	15 181	112 33	67		177 36
181	113	8 8	49	146	16 182	134 59	71		37
182	121	11 11	68	148	24	155 79	78		42
	173	31 31	75	181	33	165 103	82		59
		41 36	124	182	148	112	88		67
		44 41	140		181	165	89		71
		49 44	167		182	182	92		77
		54 49	174				98		85
		151 54	182				130		90
		161 151					134		98
		162 161					140		101
		173 162					151		106
		181 173					153		122
		181					167		130
							176		134
							177		143
							182		146
									156
									163
									167
									174

Table A11. Indices of Failing Labs
Correlation Scores Atomic Emission
Synthetic Oils August 1986

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
94	23 136	1	15 177	32 32	16 88	114	16 179	16 140	164
97	24 140	78	33	49 88	25 132	164	78	45	182
164	54 177	108	136	95 132	27 143	173	95	75	
173	88	164	164	97 141	40 160	178	114	78	
	97	175		141 160	72	182	134	87	
	114	177		151	83		135	132	
	134	182		156	88		164	157	
	136			163	93		173	164	
	155			164	94		182	181	
	164			173	95				
	167				97				
	177				114				
					135				
					147				
					151				
					164				
					168				
					173				
					175				
					176				

Table A12. Indices of Failing Labs
Proposed Scores Atomic Emission
Synthetic Oils August 1986

Fe	Ag	Al	Cr -	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
164 140	114 71	1 177	15 136	32 32	16 88	114 140	16 179	132 140	164 82
173 173	136 136	78	33 177	141 88	173 160	164 173	114		182 88
	177 140	108	164	151 132		173	134		132
	151 164			163 141		178	173		135
	173 175			164 160		182			177
	177	177							

Table A13. Indices of Failing Labs
Correlation Scores Atomic Absorption
Used Oils August 1986

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
14	12				14	3 8	9	14 14	12 12
15					15	8 11	10		19
16					16	9 24	12		25
17					17	11	24		26
24					24	12	26		31
34					26	18			
					33	22			
						23			
						24			
						28			
						31			

Table A14. Indices of Failing Labs
Proposed Scores Atomic Absorption
Used Oils August 1986

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
	12	27				11 8	26 9	14 14	12 12
		31				11	26		26
						24			

Table A15. Indices of Failing Labs
Correlation Scores Atomic Absorption
Synthetic Oils August 1986

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
8	7 28	7 7	4 31	8 12	7 7	7	7	7 14	12 12
12	9	8 31	10	12	9 32	8	8	10	13 32
24	12	10	12	18	10 34	9	12	12	19
26	17	12	17	32	12	11	13	14	23
31	24	28	18		17	16	20		25
32	28	31	24		24	18	23		32
34	33	32	25		25	22	26		33
	36		26		27	23	27		34
			29		28	24	28		35
			30		31	26	29		
			31		32	31	30		
			32		33		32		
			37		34				
					37				

Table A16. Indices of Failing Labs
Proposed Scores Atomic Absorption
Synthetic Oils August 1986

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
24	28 28	12	17 31	12 12	12 7	20	12	14	23 12
		28	24	18	37 34		26		32 32

Table A17. Indices of Failing Labs
Correlation Scores Atomic Emission
Used Oils January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
37 51	22	10 6	21	1 130	23	19 126	35 79	10 1	25 25
51	161	12 47		6 158	92	23 180	79	12 38	35 135
75		18		12 169	156	92	164	37 57	161
78		23		23 172	157	126		42 61	
108		24		26	160	152		59 135	
135		34		38		180		76 157	
156		35		42				80 160	
172		37		89				93	
176		42		93				114	
		45		100				126	
		58		101				134	
		59		117				135	
		78		124				138	
		80		128				145	
		88		130				157	
		111		133				168	
		118		135				171	
		124		136					
		132		140					
		149		141					
		154		147					
		158		154					
		160		156					
		161		160					
		168		164					
		172		167					
		174		168					
		180		172					
				173					
				174					
				176					

Table A18. Indices of Failing Labs
Proposed Scores Atomic Emission
Used Oils January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
37 17	22 45	118 6	21 32	42 130	23 1	19 126	35 5	1	25 19
51 51	161 108	158 23	63	140 158	92 5	23 180	79 6	38	35 23
108 79	109	161 47	70	172 169	156 6	92	164 13	57	161 25
135 89	168	172 101	118	176 172	157 30	126	19	61	31
172 108		158	119		160 34	152	27	135	49
176 172			128		40	180	31	157	72
			141		44		59	160	104
			144		49		64		135
			150		67		72		151
			155		75		79		156
			161		86		99		168
			169		96		140		177
			172		100		150		
			177		108		180		
					115				
					118				
					122				
					124				
					140				
					146				
					156				
					160				

Table A19. Indices of Failing Labs
Correlation Scores Atomic Emission
Synthetic Oils January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
108 50	4 37	10 71	154 74	23	2 102	29	12 51	114 80	12
128 102	23	25 108	157 108	118	12 103	63	31 102	160 89	170
170 108	31	41 110	119	154	31 108	68	38 108	135	
171	34	76 124	174	173	48 138	104	71 120	147	
173	38	101 158		178	54	108	77	172	
178	71	102 174			65	130	79		
	75	138 180			79	170	103		
	76	143			108	173	108		
	102	161			115		137		
	135	170			127		154		
	148	178			148		155		
	151	180			159		159		
	152				168		161		
	154				171		164		
	157				175		170		
	161				178		172		
	172						173		
	176						175		
	178						178		

Table A20. Indices of Failing Labs
Proposed Scores Atomic Emission
Synthetic Oils January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
108 50	31 37	138 71	154 74	23	31 102	29 29	79 108	114 80	12 17
128 67	151 176	161 108	157 108		108 103	63 42	154	160 89	170 71
170 102	176	110	119		115 108	68 63	161	135	79
171 108		124	174		171	104 68	164	147	156
178 136		158			175	108 136	173	172	
		174			178	170 173	178		
		180				173			

Table A21. Indices of Failing Labs
Correlation Scores Atomic Absorption
Used Oils January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
19		19	36	9 33	20	9 9	7	7	
28		37			39	10 14	10	20	
29						13	23		
37						14			
						20			
						23			

Table A22. Indices of Failing Labs
Proposed Scores Atomic Absorption
Used Oils January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
28 33	29	26	11 3	33	39	13 14	7 7	7 7	24
37			36 11			14	23 10	20 20	
							23		

Table A23. Indices of Failing Labs
Correlation Scores
Synthetic Oils
Atomic Absorption
January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
9 11	3	11	3 33	26 11	9 3	13 20	4 9	20	16 16
11 36	5	19	11	38 36	10	20	7 17		28 28
25 37	11	31	19		11		9		31
29	15	32	22		28		10		34
35	19		26		31		16		
36	34		28		35		17		
37	35		31		36		21		
	38		33		39		23		
			35				25		
			36				27		
			37						
			38						

Table A24. Indices of Failing Labs
Proposed Scores
Synthetic Oils
Atomic Absorption
January 1985

Fe	Ag	Al	Cr	Cu	Mg	Si	Ti	Mo	Ni
Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep	Acc Rep
9	19	11 11		11	9	13	25	20 20	31 28
29		19		36	35	20			
36									

APPENDIX B

This appendix describes the mathematical computations used for the proposed scoring system. The same procedure is followed for each element, for each type of sample, for either type of instrument. As in the text above, let

x_i = Analysis result for sample 1, laboratory i

y_i = Analysis result for sample 2, laboratory i.

The first step is to perform the bivariate trimming. Define

$\bar{x} = \Sigma x_i / n$ = Average of all sample 1 results

$\bar{y} = \Sigma y_i / n$ = Average of all sample 2 results

where n is the total number of instruments analyzing this sample. Now define

$$A = \Sigma (x_i - \bar{x})^2 / (n - 1)$$

$$C = \Sigma (y_i - \bar{y})^2 / (n - 1)$$

$$B = \Sigma (x_i - \bar{x})(y_i - \bar{y}) / (n - 1)$$

and let S be the 2 by 2 matrix whose first row is A, B and whose second row is B, C; define T to be the matrix inverse of S. This matrix T is used to evaluate n quadratic forms, one for each participating laboratory. That is, for instrument i the quadratic form is

$$Q_i = t_{11}(x_i - \bar{x})^2 + 2t_{12}(x_i - \bar{x})(y_i - \bar{y}) + t_{22}(y_i - \bar{y})^2$$

where the first row of T is t_{11} , t_{12} and the second row of T is t_{12} , t_{22} . These Q_i values then are ranked in order of magnitude, from smallest to largest and are used to trim off (no more than) 5% of the instruments; if for example $n = 183$ instruments had analyzed the sample, 5% of n equals 9.15 so the 9 largest Q_i values identify those instruments to be trimmed off. The remaining 174 laboratories are used to determine the accuracy and repeatability limits. Let m represent the number of instruments remaining after trimming; m, of course, is the next larger integer above .95n (or .95n rounded up).

The trimmed means, for the x and y scores, are the averages of the m remaining pairs. Hopefully without confusion, let \bar{x} and \bar{y} represent these trimmed means and define

$$X_i = x_i - \bar{x}$$

$$Y_i = y_i - \bar{y}$$

for all n instruments. The accuracy score for instrument i then is

$$A_i = (X_i + Y_i)/2^{.5}$$

and the repeatability score for instrument i is

$$R_i = (X_i - Y_i)/2^{.5}.$$

We now have n pairs, (A_i, R_i) , one for each instrument. Using m instrument pairs which were not trimmed initially, define the accuracy and repeatability standard deviations by

$$s_a = (\Sigma A_i^2 / (m - 1))^{.5}$$

$$s_r = (\Sigma R_i^2 / (m - 1))^{.5}.$$

The accuracy limit is $3s_a$ and the repeatability limit is $3s_r$; any instrument, trimmed or not, whose accuracy score or repeatability score exceeds the respective limit in magnitude (absolute value) fails on that score.

DISTRIBUTION LIST

	<u>NO. OF COPIES</u>
Library (Code 0142) Naval Postgraduate School Monterey, CA 93943-5000	2
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Office of Research Administration (Code 012) Naval Postgraduate School Monterey, CA 93943-5000	1
Center for Naval Analysis 4401 Ford Ave. Alexandria, VA 22302-0268	1
Library (Code 55) Naval Postgraduate School Monterey, CA 93943-5000	1
Operations Research Center, Rm E40-164 Massachusetts Institute of Technology Attn: R. C. Larson and J. F. Shapiro Cambridge, MA 02139	1
Koh Peng Kong OA Branch, DSO Ministry of Defense Blk 29 Middlesex Road SINGAPORE 1024	1
Arthur P. Hurter, Jr. Professor and Chairman Dept of Industrial Engineering and Management Sciences Northwestern University Evanston, IL 60201-9990	1
Institute for Defense Analysis 1800 North Beauregard Alexandria, CA 22311	1
Kelly Air Force Base TX 79241	1

DUDLEY KNOX LIBRARY



3 2768 00329691 4